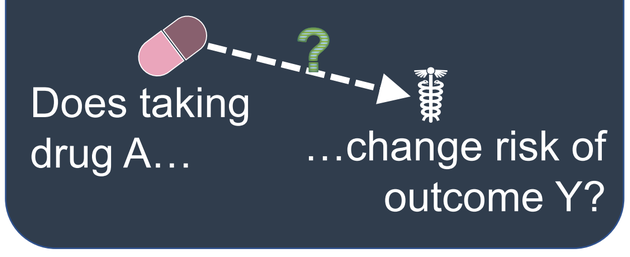
Now published in *JAMIA Open* https://doi.org/10.1093/jamiaopen/ooaa040

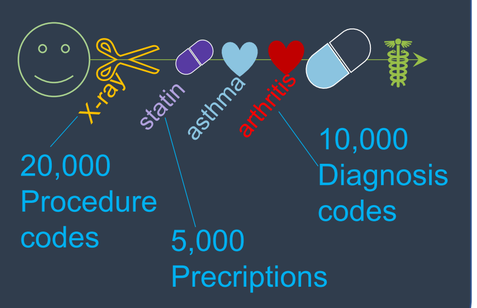
Capturing medical practice with indication embeddings

Rachel Melamed
Biology @ UMass Lowell
Rachel_Melamed@uml.edu
https://rdmelamed.github.io/
@rdmelamed

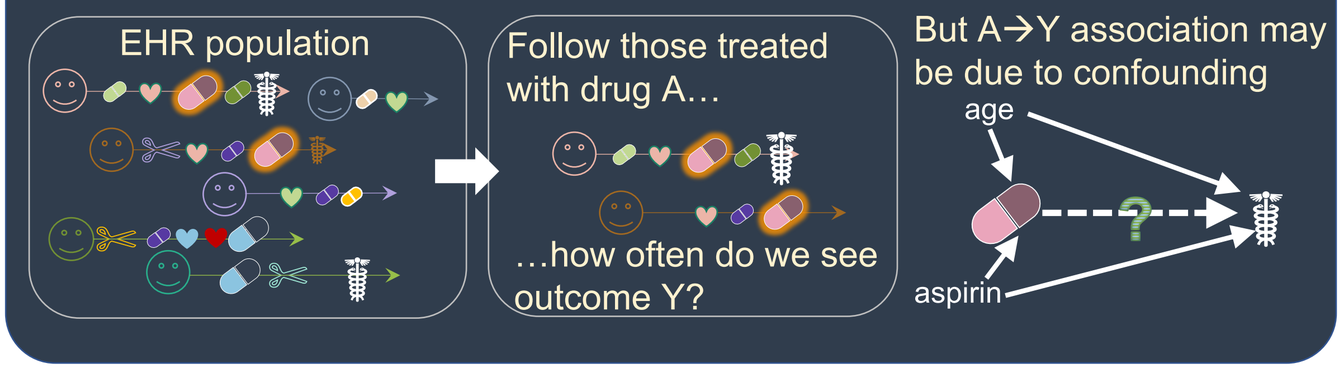
Drug safety & comparative effectiveness



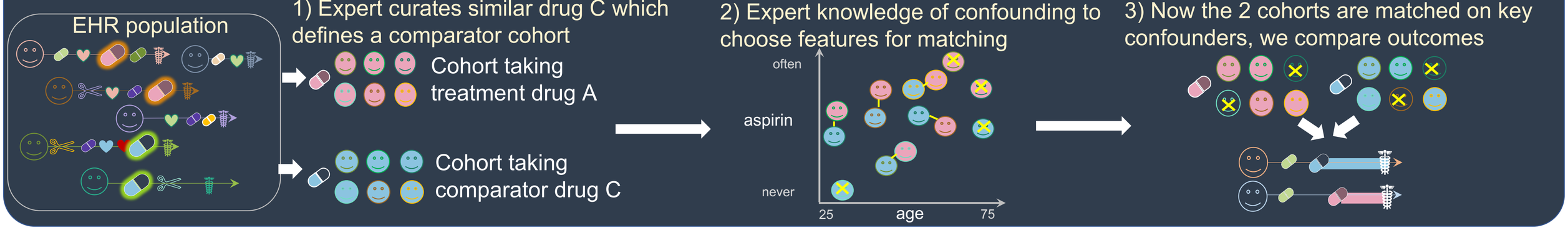
Health data



Goal: health data to find drug effects



Cohort studies reduce confounding by crafting comparator cohort

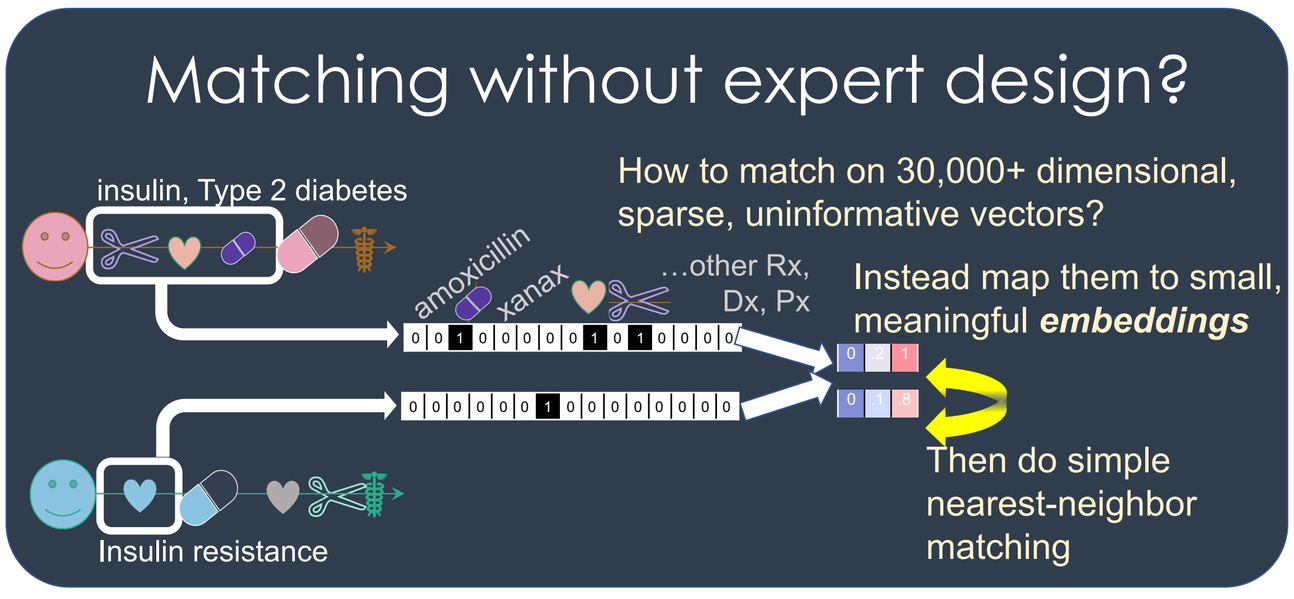


Cohort study relies on expert design

Domain experts needed to:

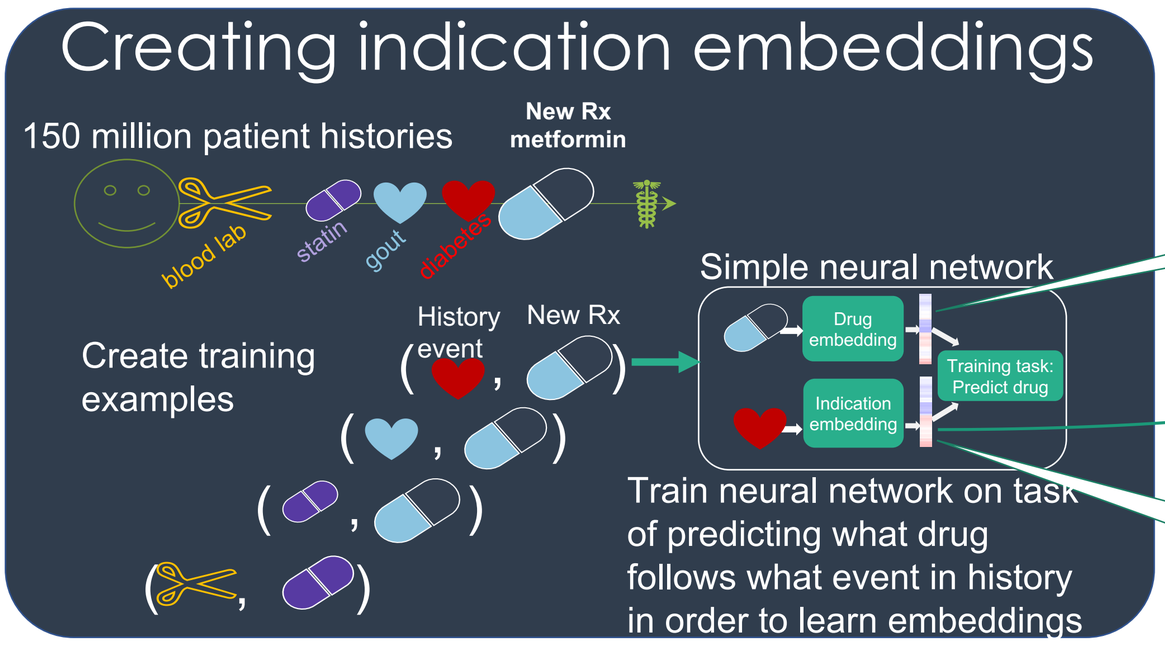
- **Expert Task 1**: Find suitable comparator drug
- **Expert Task 2**: Design matching—identify confounders

Matching without expert design?



Medical knowledge created the health data: can we extract this meaning from the health data to create embeddings?

Creating indication embeddings

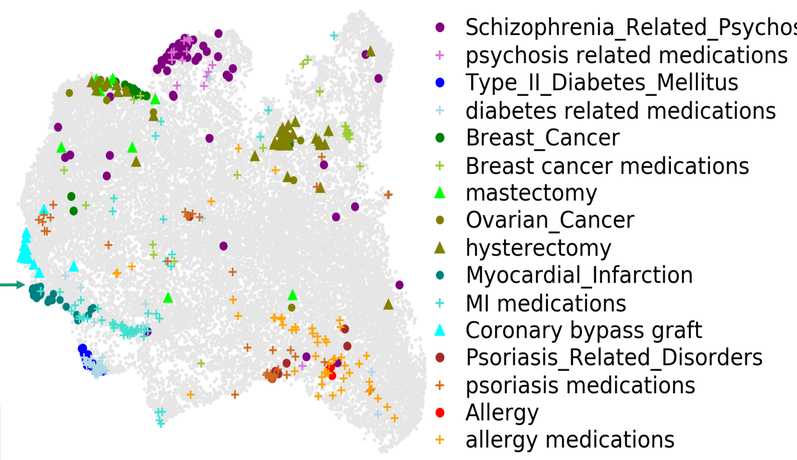


Embeddings relate codes to health needs

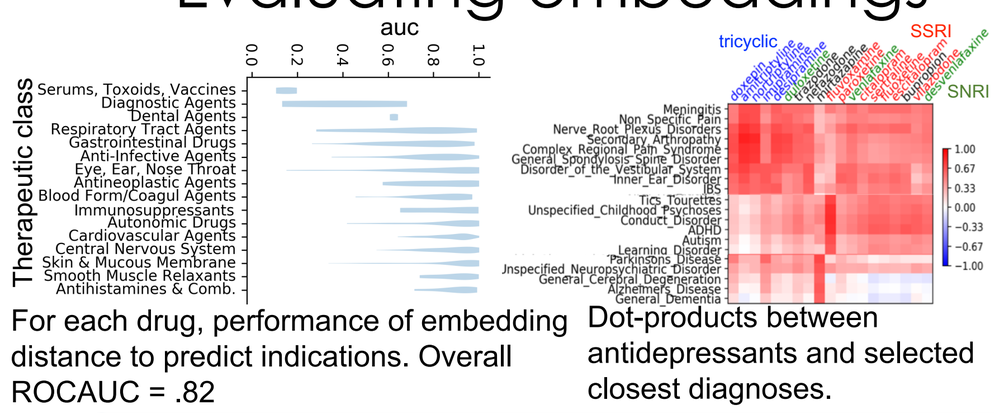
Drug embedding = drugs given in most similar health contexts

Map each event to 50-dimensional vector

Indication embedding = health context for prescription of a new drug



Evaluating embeddings



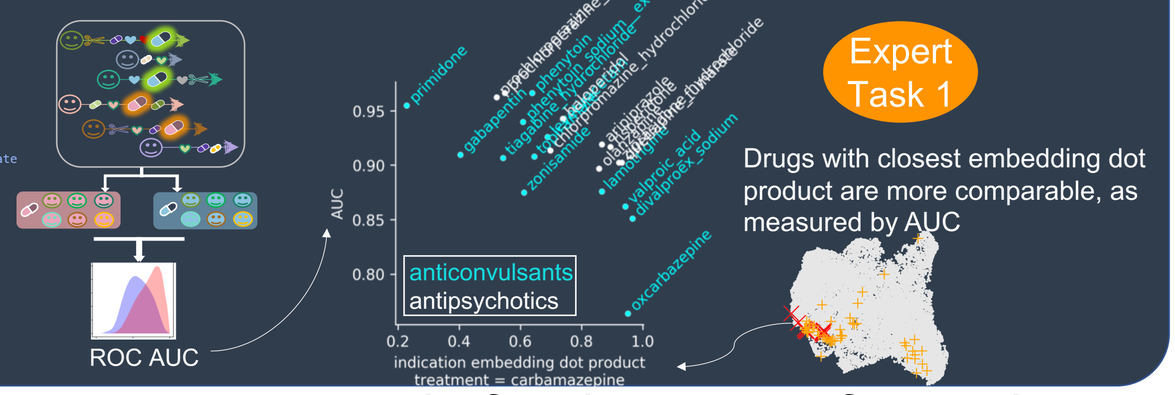
For each drug, performance of embedding distance to predict indications. Overall ROCAUC = .82

Dot-products between antidepressants and selected closest diagnoses.

Embeddings identify drugs with similar users

Drugs with same therapeutic use as **carbamazepine**: primarily anticonvulsant, off-label for bipolar.

aripiprazole	oxcarbazepine
asenapine	paliperidone
carbamazepine	perphenazine
chlorpromazine	primidone
clozapine	prochlorperazine
divalproex_sodium	prochlorperazine_maleate
fluphenazine	quetiapine_fumarate
gabapentin	risperidone
haloperidol	thioridazine
haloperidol_lactate	thiothixene
lacosamide	tiagabine
lamotrigine	topiramate
levetiracetam	trifluoperazine
lithium_carbonate	valproic_acid
lurasidone	ziprasidone
Olanzapine	zonisamide

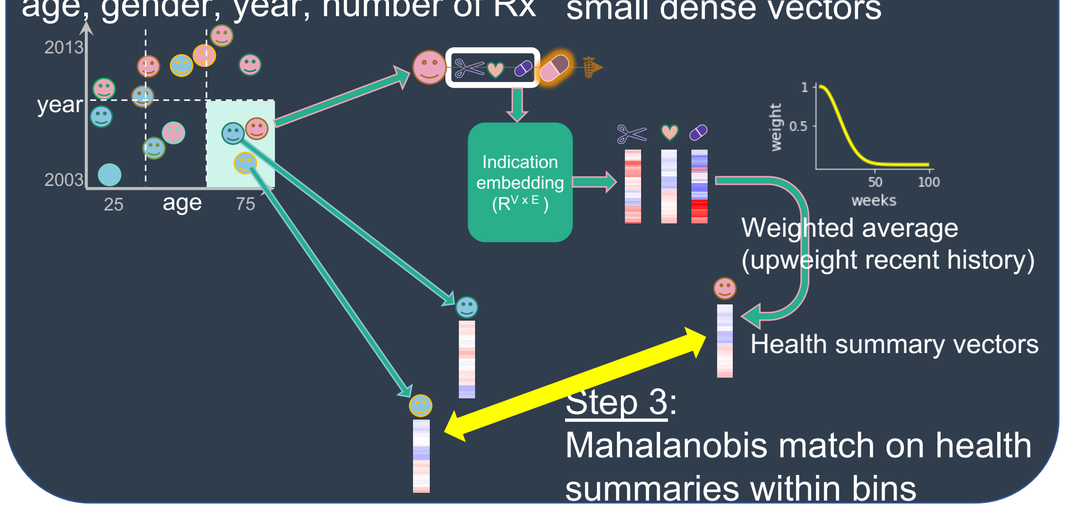


Match with embeddings

Expert Task 2

Step 1: Coarsened exact matching by age, gender, year, number of Rx

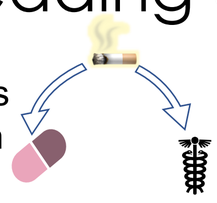
Step 2: Encode histories → small dense vectors



Step 3: Mahalanobis match on health summaries within bins

Embedding can match for key confounders

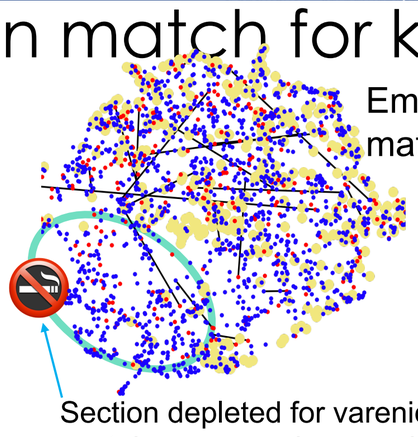
Smoking confounds bupropion effects.



Each point is UMAP placement of embedding of one patient taking either:

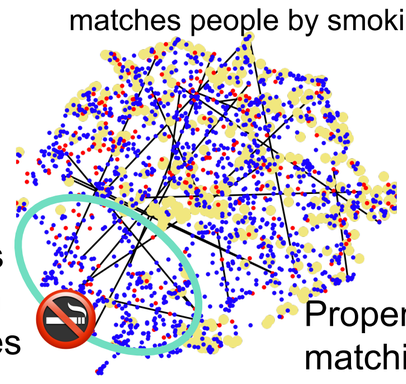
- bupropion (antidepressant also prescribed for smoking cessation)
- escitalopram (another antidepressant)
- varenicline (smoking cessation drug)

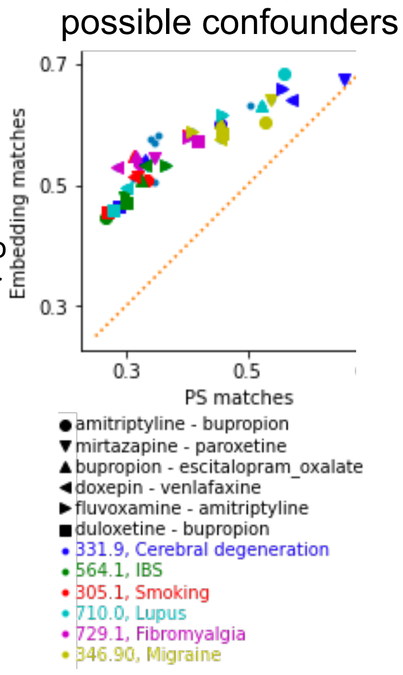
Lines connect matched pairs of bupropion to escitalopram takers in 2 matching schemes



Embedding matching

Section depleted for varenicline users so contains non-smokers: embedding better matches people by smoking status



Propensity score matching

Embedding matches better for many possible confounders